|  |  |
| --- | --- |
|  |  |
| TIME SERIES FORECASTING PROJECT |  |
|  |  |
|  | Submitted by,VIDYA V |
|  | PGPDSBA.O.2023.B 11.11.2023 |

**CONTENTS**

|  |  |
| --- | --- |
| Sparkling Wine Sales Dataset | 4 |
| 1. Read the data as an appropriate Time Series data and plot the data. | 4 |
| 1. Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition. | 5 |
| 1. Split the data into training and test. The test data should start in 1991. | 8 |
| 1. Build all the exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other additional models such as regression, naïve forecast models, simple average models, and moving average models should also be built on the training data and check the performance on the test data using RMSE. | 8 |
| 1. Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment. Note: Stationarity should be checked at alpha = 0.05. | 12 |
| 1. Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE. | 14 |
| 1. Build a table with all the models built along with their corresponding parameters and the respective RMSE values on the test data. | 17 |
| 1. Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands. | 18 |
| 1. Comment on the model thus built, report your findings, and suggest the measures that the company should be taking for future sales. | 19 |
|  |  |

**List of Figures**

|  |  |
| --- | --- |
| **Name** | **Page No.** |
| Fig.1.1. Sparkling Dataset | 4 |
| Fig.1.2. Sparkling Dataset Description | 4 |
| Fig.1.3. Sparkling wine sales | 5 |
| Fig.1.4 Yearly boxplot- Sparkling Dataset | 5 |
| Fig.1.5. Monthly Boxplot- Sparkling wine sales | 6 |
| Fig.1.6. Month-plot- Sparkling wine sales | 6 |
| Fig.1.7. Monthly sales over the years | 6 |
| Fig.1.8. Yearly sales- Sparkling Wine | 7 |
| Fig.1.9. Additive Decomposition- Sparkling sales | 7 |
| Fig.1.10. Multiplicative Decomposition- Sparkling wine sales | 8 |
| Fig.1.11 Linear Regression model test forecast plot- Sparkling sales | 8 |
| Fig.1.12. Naive forecast of test data- Sparkling wine sales | 9 |
| Fig.1.13. Simple Average forecast of test data- Sparkling wine sales | 9 |
| Fig.1.14.Moving Average forecast of test data- Sparkling wine sales | 10 |
| Fig.1.15**.** Simple Exponential smoothing forecast of test data- Sparkling wine sales | 10 |
| Fig.1.16. Simple Exponential smoothing forecast of test data- Sparkling wine sales optimized for lowest RMSE | 11 |
| Fig.1.17**.** Holt forecast of test data- Sparkling wine sales | 11 |
| Fig.1.18. Holt forecast of test data- Sparkling wine sales - optimized for lowest RMSE | 11 |
| Fig.1.19**.** Holt Winters forecast of test data- Sparkling wine sales | 12 |
| Fig.1.20. Holt Winters smoothing forecast of test data- Sparkling wine sales - optimized for lowest RMSE | 12 |
| Fig.1.21. Sparkling dataset -Stationarity test rolling mean and Standard deviation plots | 13 |
| Fig.1.22. Dickey Fuller Test results- Sparkling Dataset | 13 |
| Fig.1.23. Differenced series -Stationarity test rolling mean and Standard deviation plots | 14 |
| Fig.1.24. Dickey Fuller Test results | 14 |
| Fig.1.25. ARIMA results Summary- Sparkling Dataset | 15 |
| Fig.1.26. ARIMA model forecast on test data- Sparkling Dataset | 15 |
| Fig.1.27. SARIMA results Summary- Sparkling Dataset | 16 |
| Fig.1.28. SARIMA model forecast on test data- Sparkling Dataset | 16 |
| Fig.1.29. Diagnostics and Correlogram- SARIMA Model | 17 |
| Fig.1.30. Sparkling Dataset model Results- Test RMSE | 17 |
| Fig.1.31. SARIMA Model forecast for next 12 months- Sparkling Dataset | 18 |
| Fig.1.32. Holt Winters Model forecast for next 12 months- Sparkling Dataset | 18 |
| Fig.1.33. Forecast for 12 months- SARIMA and HW | 19 |

# Sparkling Dataset

For this particular assignment, the data of different types of wine sales in the 20th century is to be analysed. Both of these data are from the same company but of different wines. As an analyst in the ABC Estate Wines, you are tasked to analyse and forecast Wine Sales in the 20th century.

1. **Read the data as an appropriate Time Series data and plot the data.**

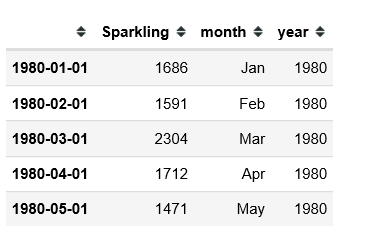


Fig.1.1. Sparkling Dataset



Fig.1.2. Sparkling Dataset Description

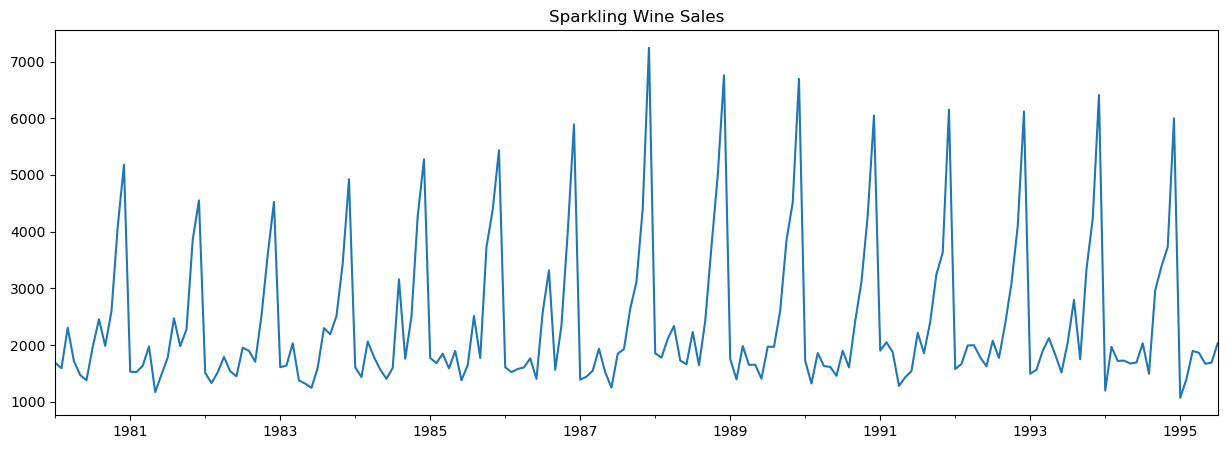


Fig.1.3. Sparkling wine sales

**Observations:**

* The plot represents the Sparkling wine sales from Jan 1980 to July 1995, covering a span of 15.5 years- 187 values
* There seems to be some seasonality associated with this plot.
* The minimum sales was 1070, the maximum sales was 7242, with a mean of 2402
* There are no null values

1. **Perform appropriate Exploratory Data Analysis to understand the data and also perform decomposition.**
   1. **EDA**

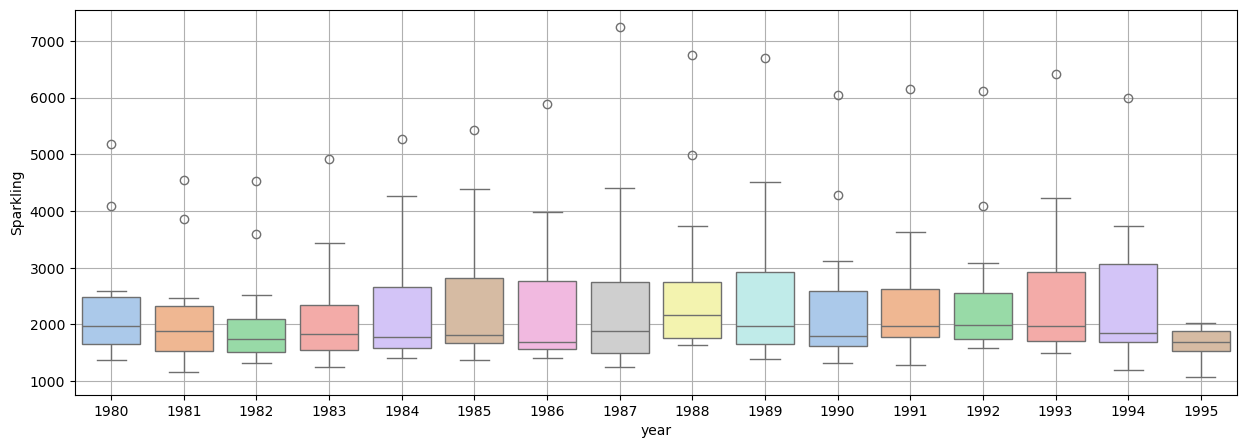


Fig.1.4 Yearly boxplot- Sparkling Dataset

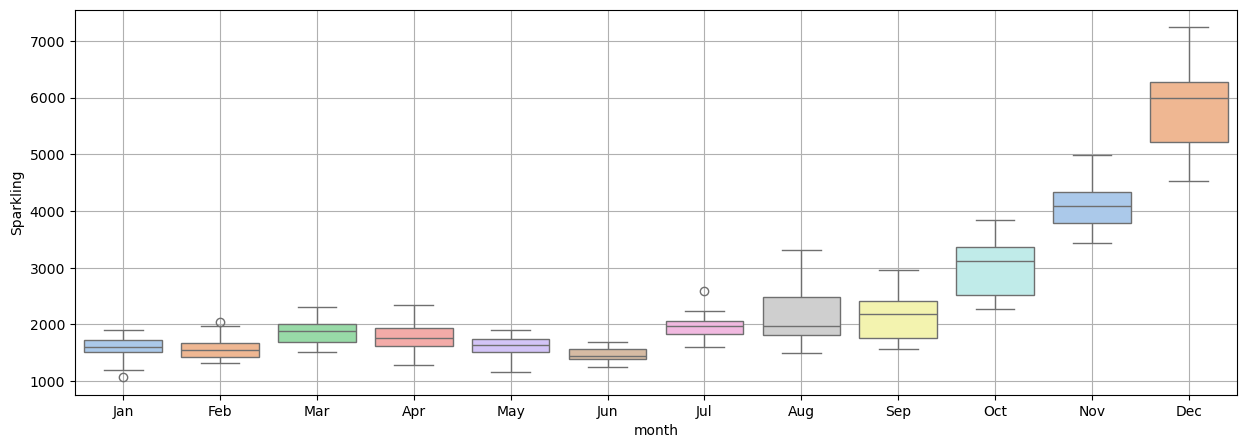


Fig.1.5. Monthly Boxplot- Sparkling wine sales

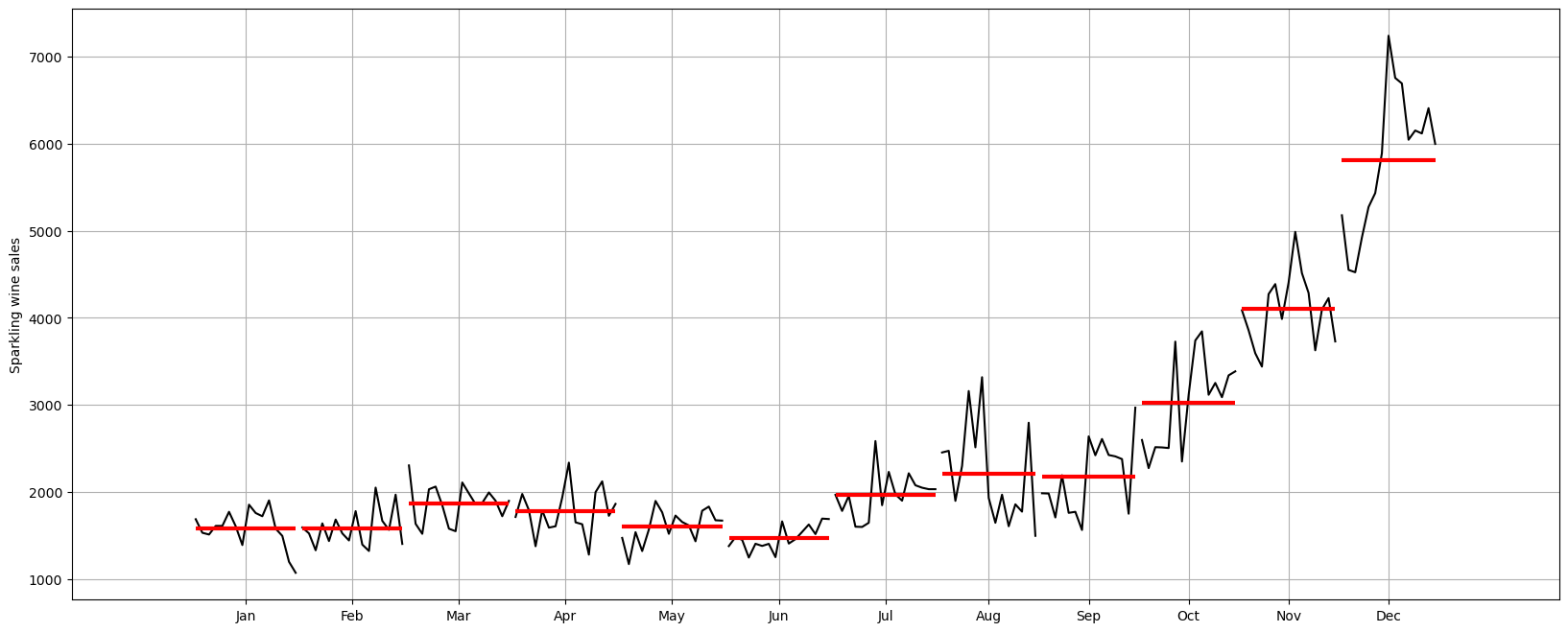


Fig.1.6. Month-plot- Sparkling wine sales

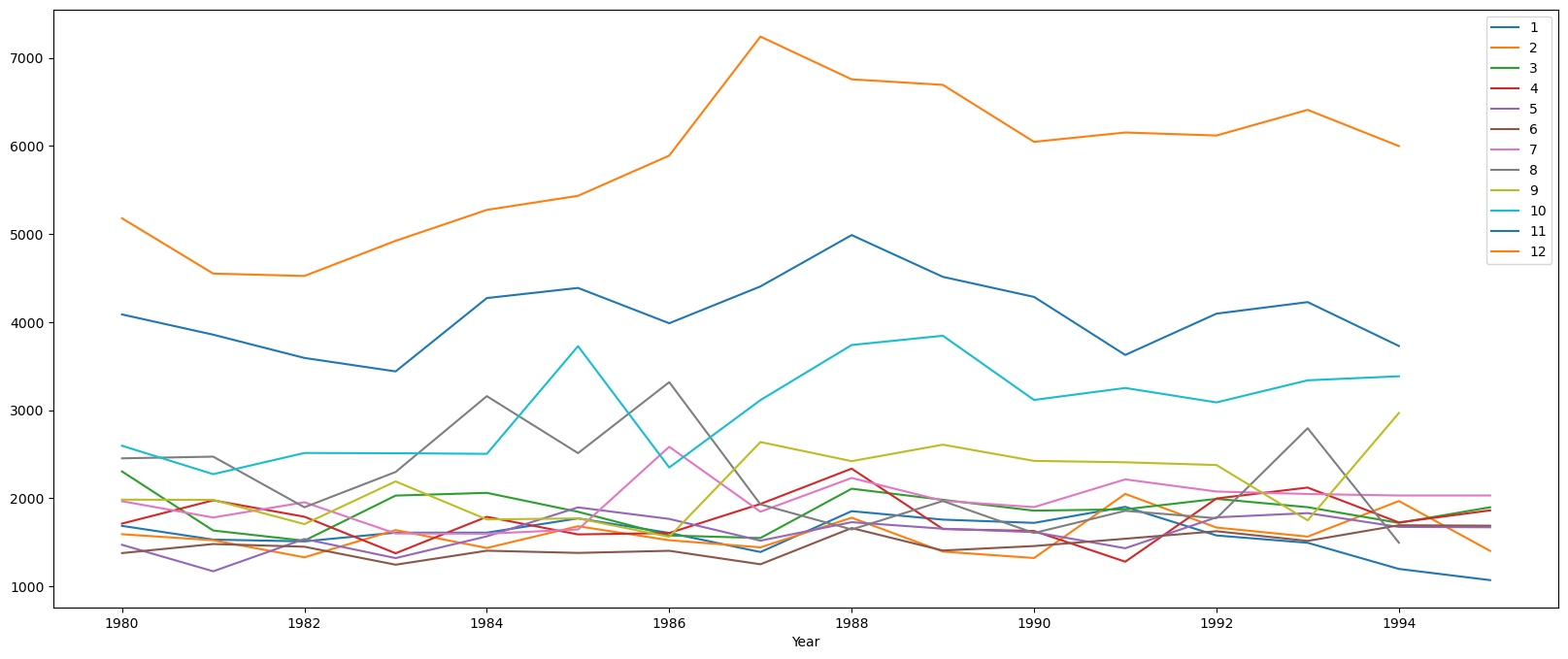


Fig.1.7. Monthly sales over the years

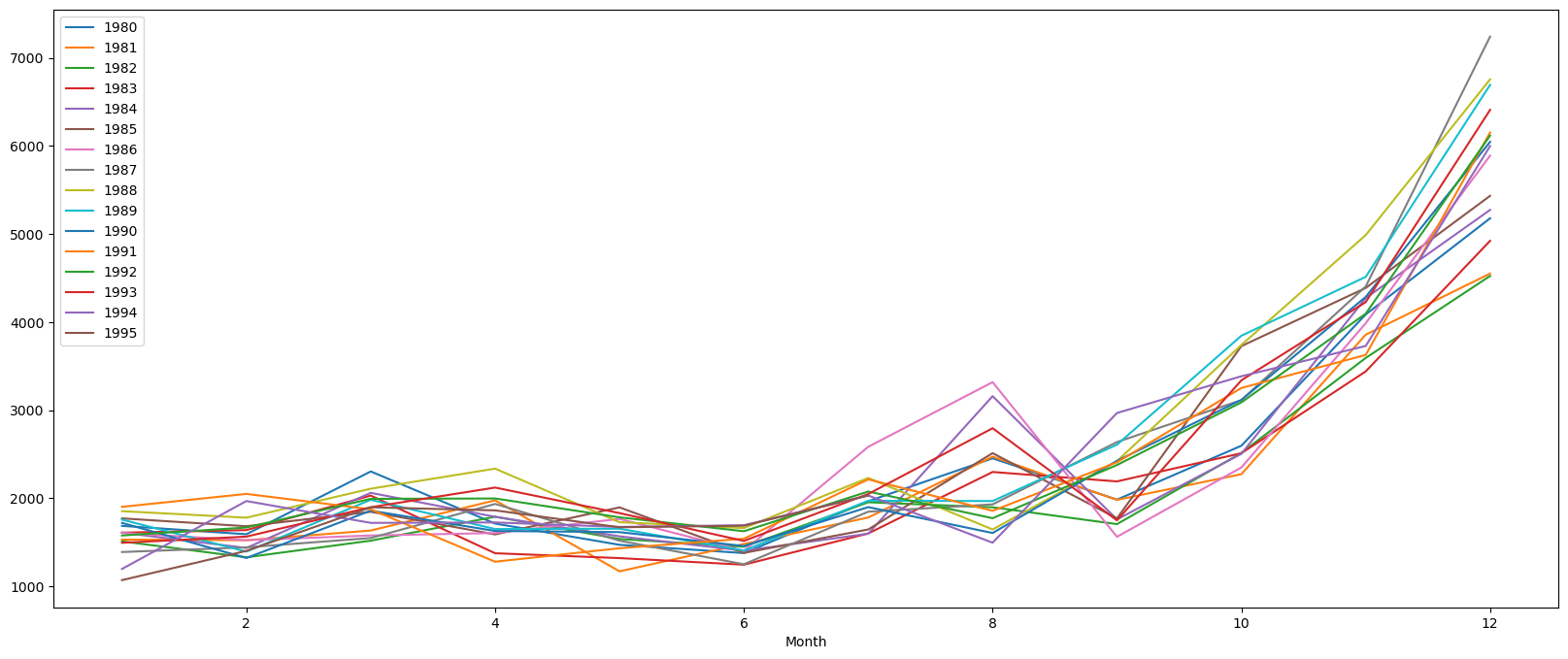


Fig.1.8. Yearly sales- Sparkling Wine

**Observations:**

* The sales remains low for the first half of the year, and increases in the second half
* Peak sales is in the month of December
* The variability also changes from Jan to Dec. Seasonality is indicated
* No trend can be discerned from the plots
  1. **Decomposition**

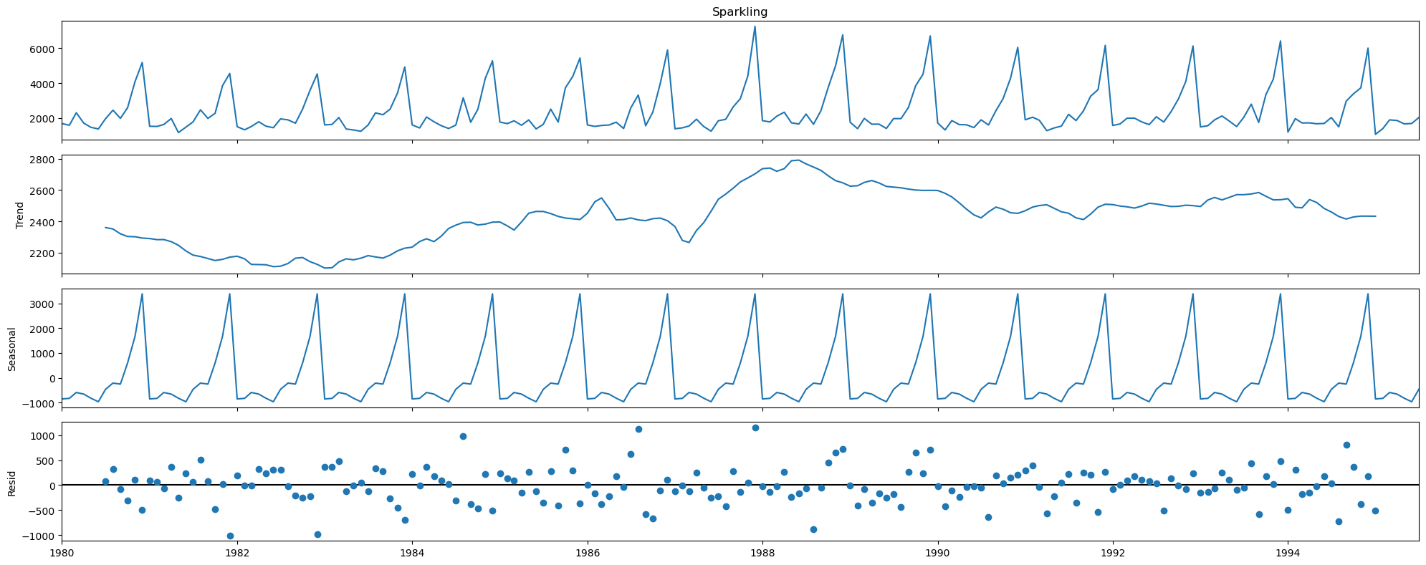
****

Fig.1.9. Additive Decomposition- Sparkling sales

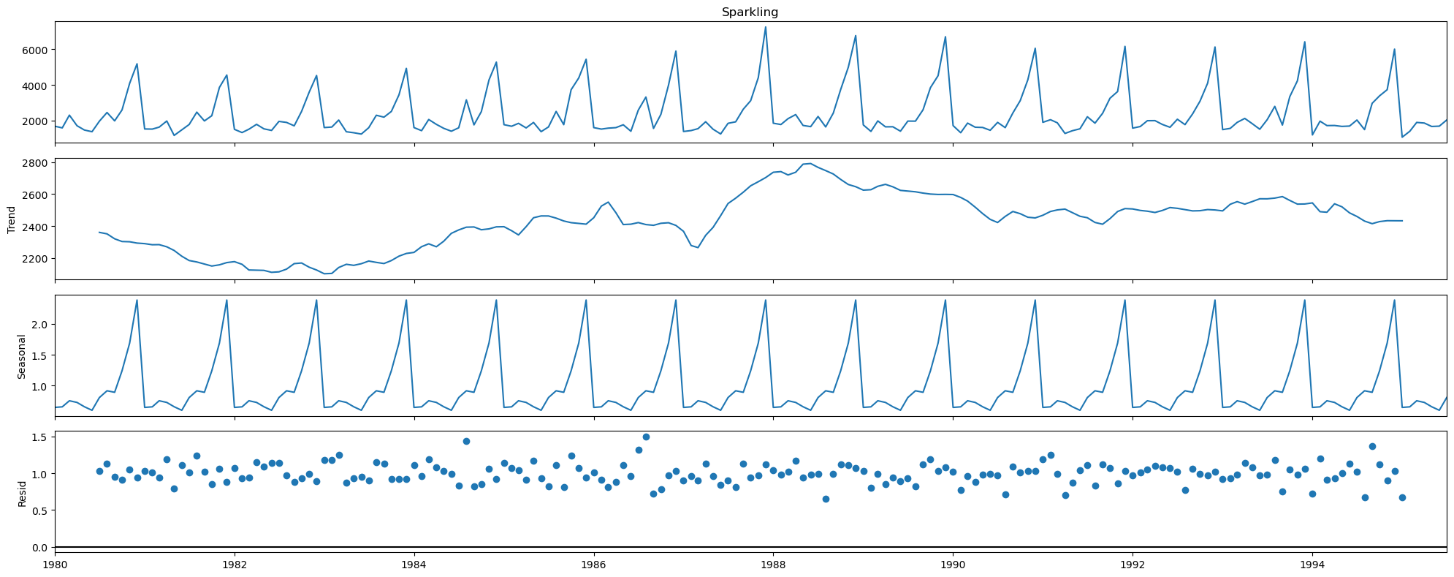


Fig.1.10. Multiplicative Decomposition- Sparkling wine sales

**Observations:**

* Clear seasonality component observed in both types of decomposition
* The residual plots of both the decompositions look similar. Hence, we can adopt the simpler of the two- additive seasonality.

1. **Split the data into training and test. The test data should start in 1991.**

**Observations:**

* After the split, the train dataset contains 132 values
* The test dataset contains 55 values, starting from Jan 1991

1. **Build all the exponential smoothing models on the training data and evaluate the model using RMSE on the test data. Other models such as regression,naïve forecast models and simple average models. should also be built on the training data and check the performance on the test data using RMSE.**
   1. **Simple Models**
      1. **Linear Regression**

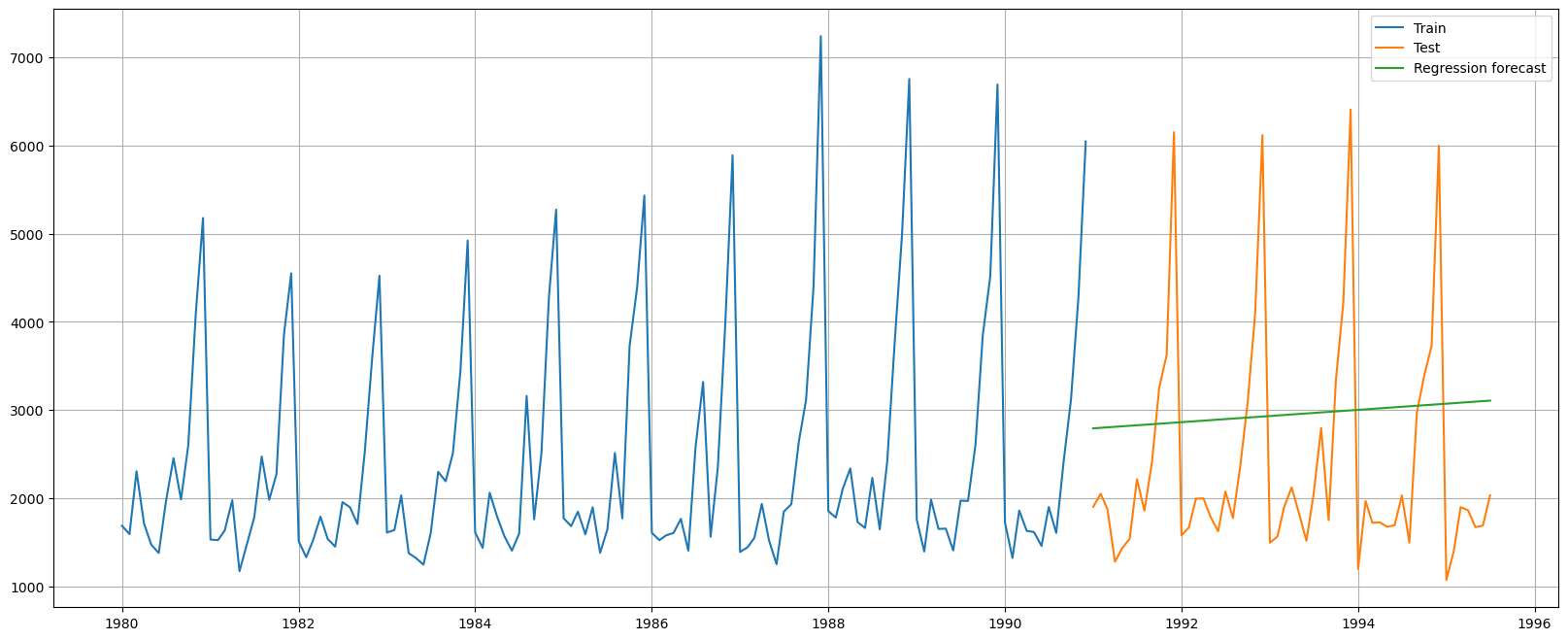
****

Fig.1.11 Linear Regression model test forecast plot- Sparkling sales

* + 1. **Naïve Forecast**

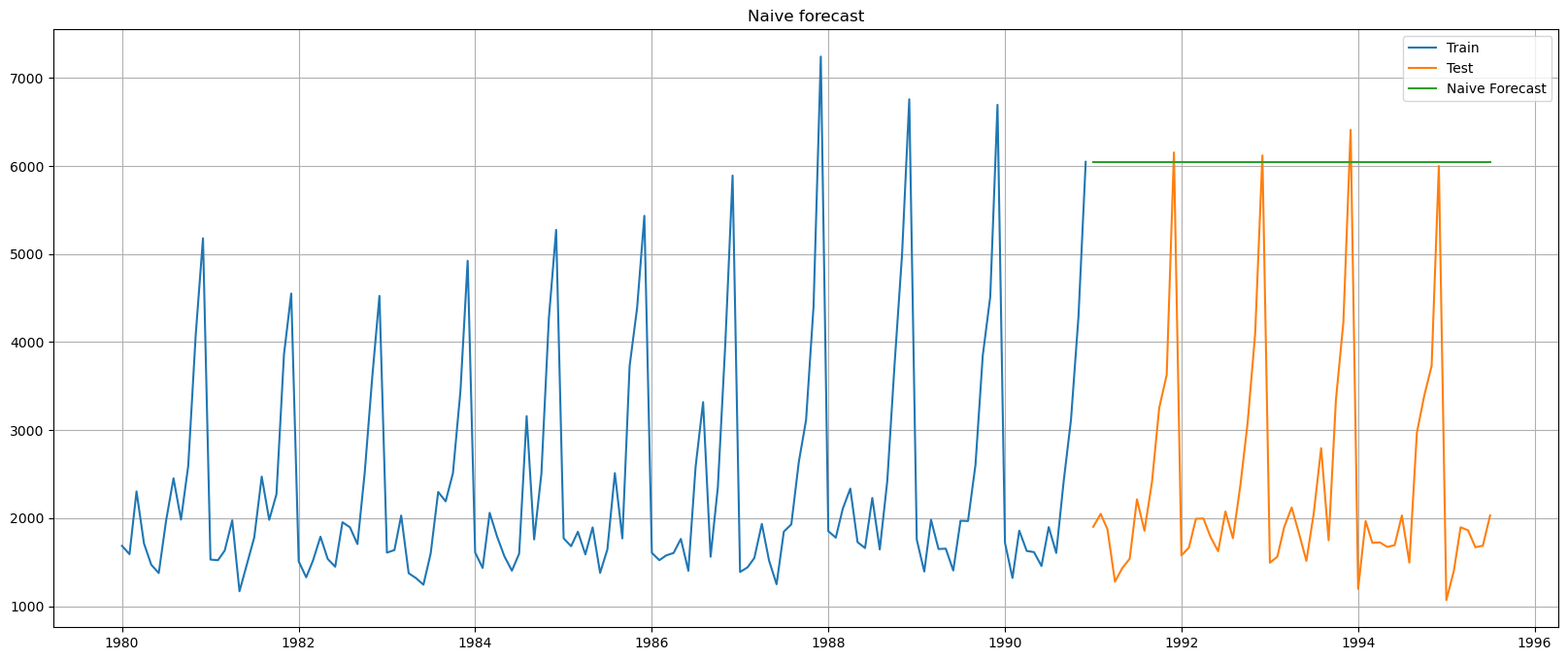
****

Fig.1.12. Naive forecast of test data- Sparkling wine sales

* + 1. **Simple Average**

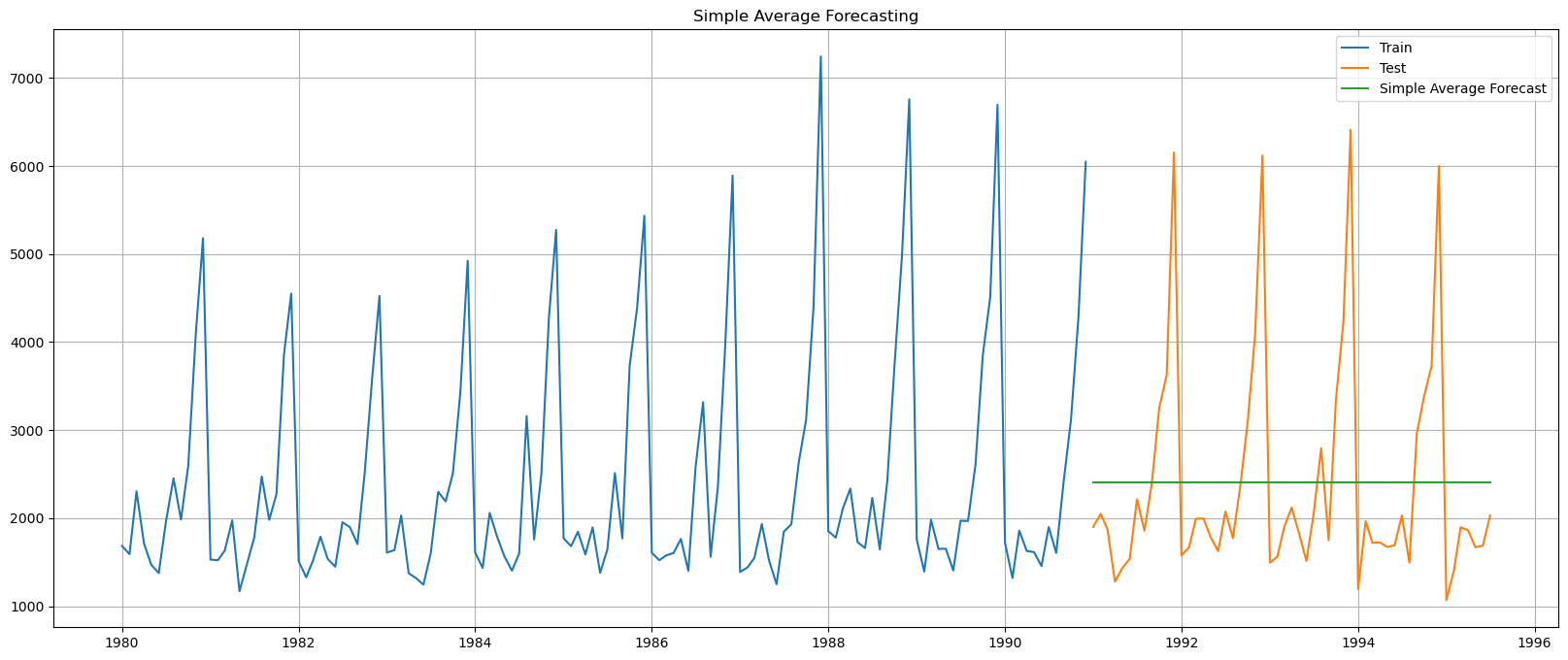
****

Fig.1.13. Simple Average forecast of test data- Sparkling wine sales

* + 1. **Moving Average**

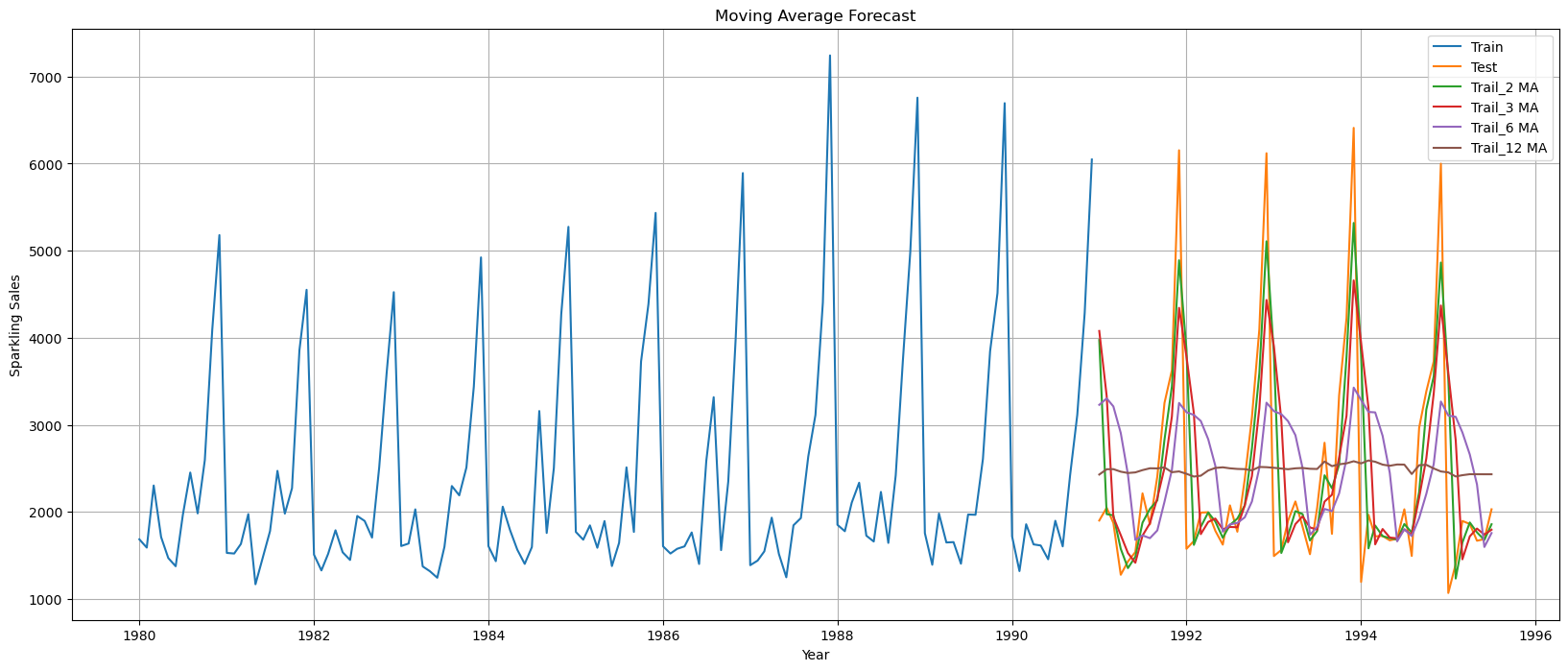
****

Fig.1.14.Moving Average forecast of test data- Sparkling wine sales

Observations:

* + - Best fit occurs in MA trail 2 model
  1. **Exponential Smoothing Models**
     1. **Simple Exponential Smoothing**

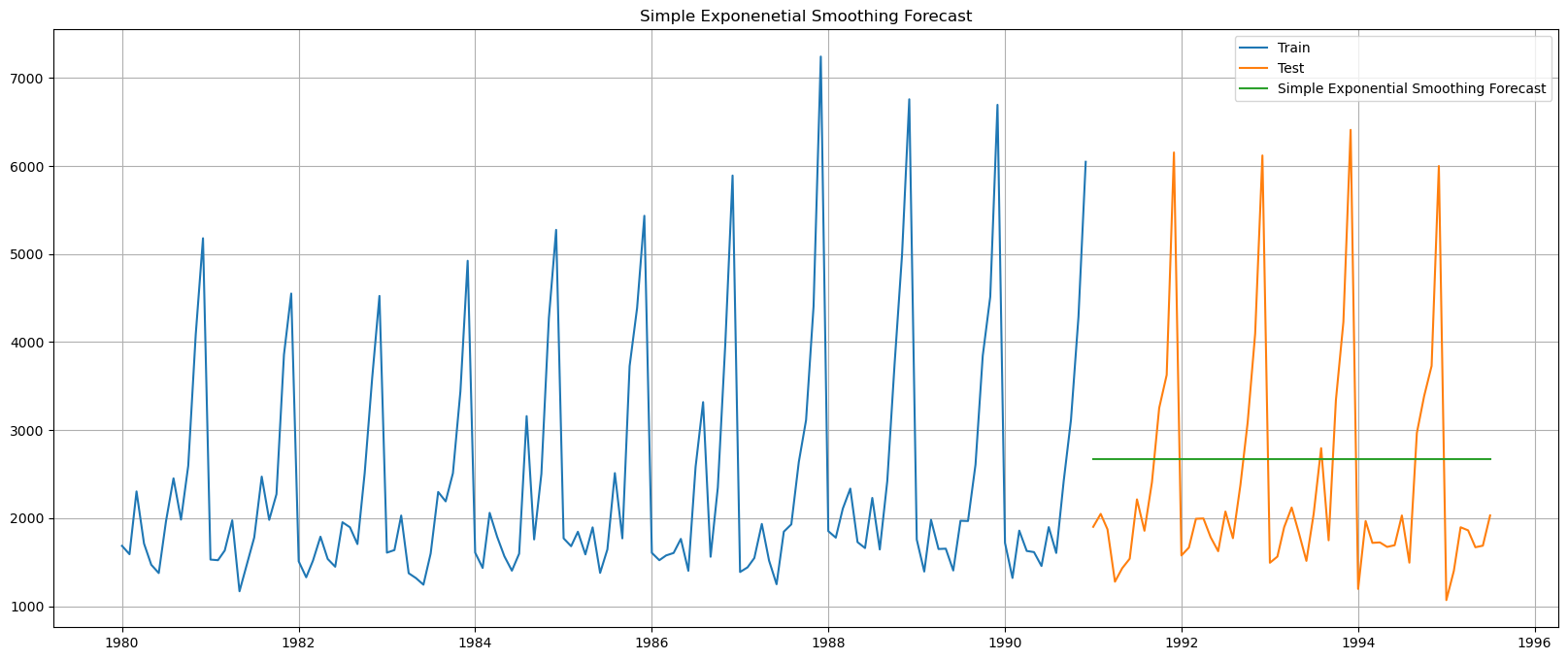
****

Fig.1.15**.** Simple Exponential smoothing forecast of test data- Sparkling wine sales

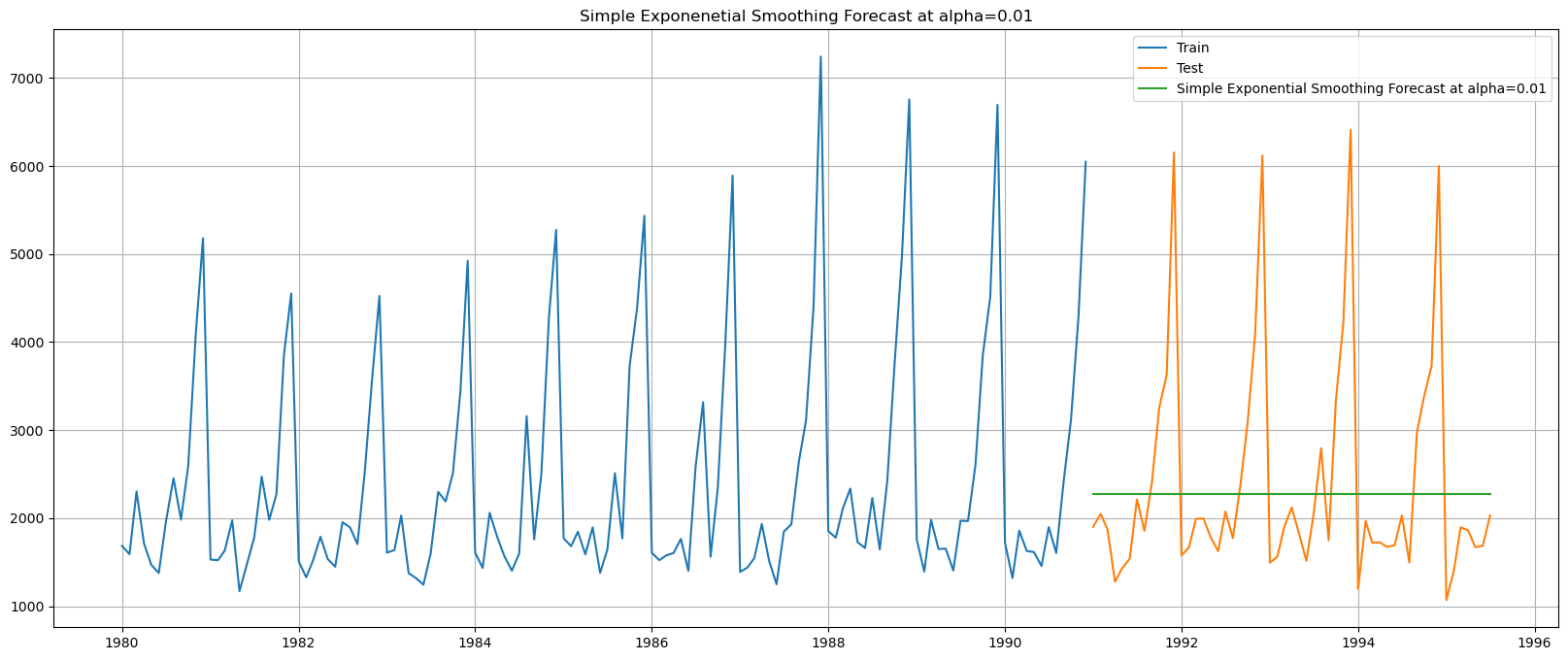
****

Fig.1.16. Simple Exponential smoothing forecast of test data- Sparkling wine sales optimized for lowest RMSE

**Observations:**

* + - RMSE is the lowest for alpha=0.01
    1. **Holt Double Exponential Smoothing**

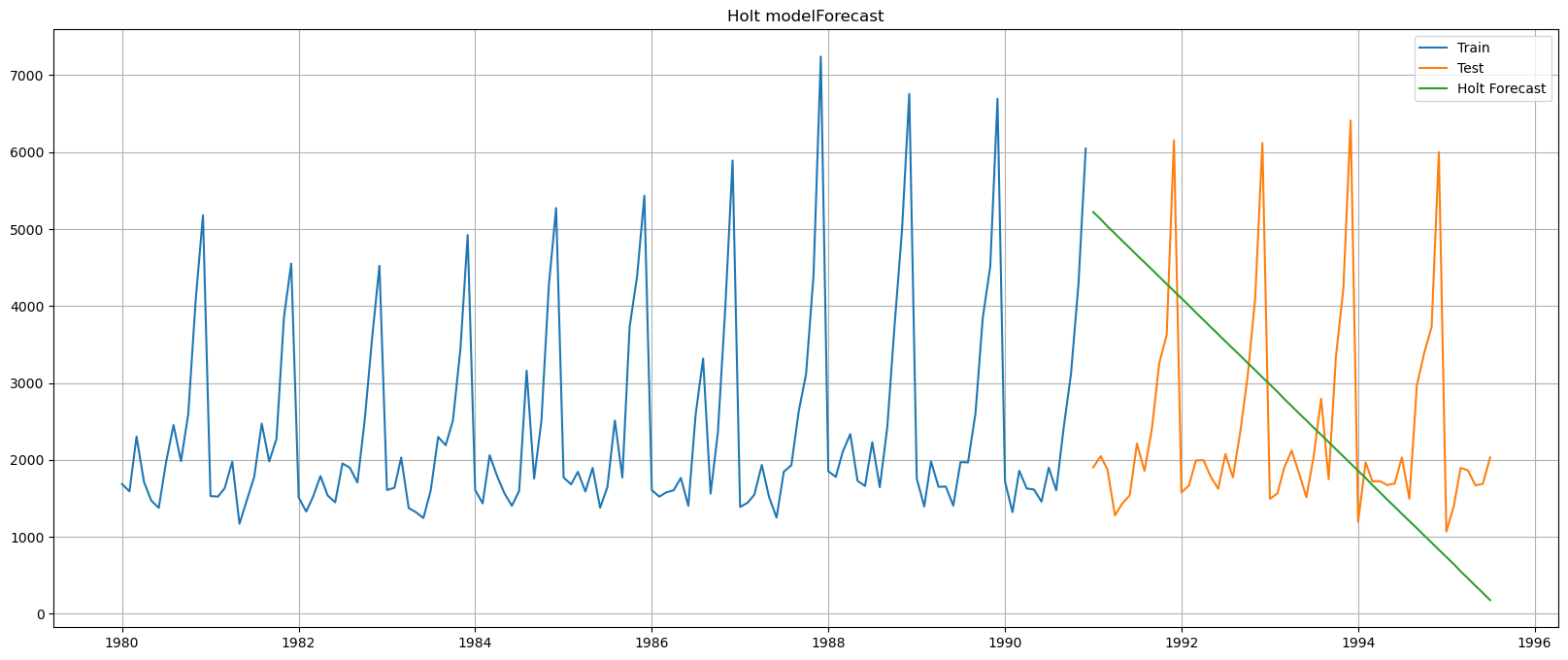
****

Fig.1.17**.** Holt forecast of test data- Sparkling wine sales

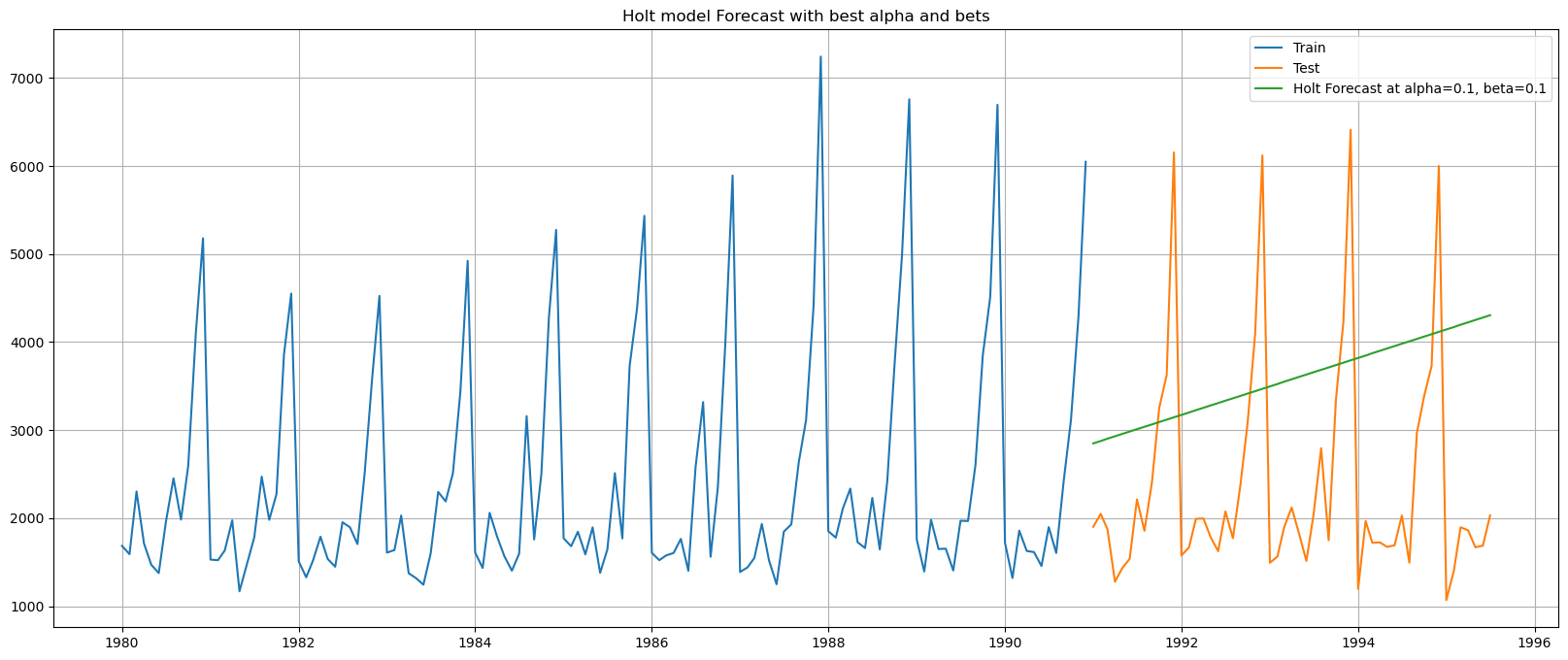
****

Fig.1.18. Holt forecast of test data- Sparkling wine sales - optimized for lowest RMSE

**Observations:**

* + - RMSE is the lowest for alpha=0.1, beta=0.1
    1. **Holt-Winters Triple Exponential Smoothing**

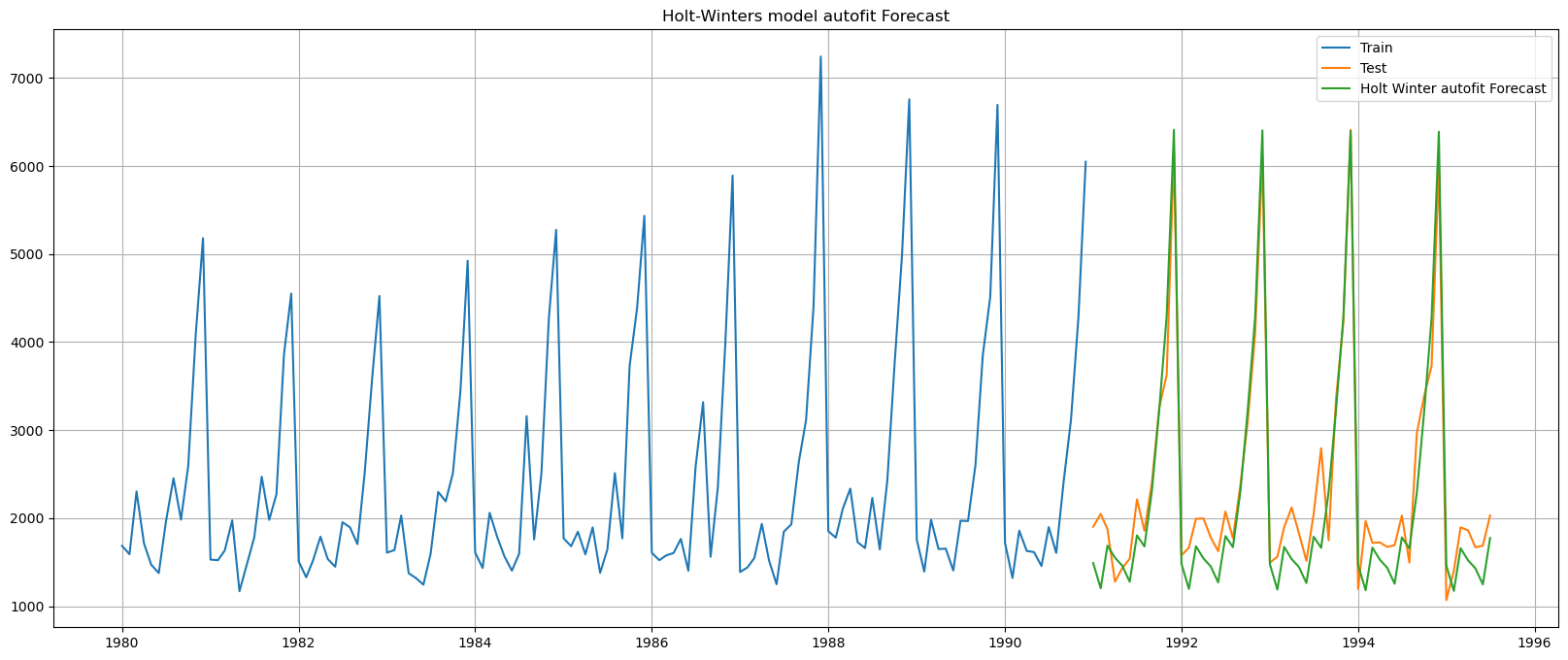
****

Fig.1.19**.** Holt Winters forecast of test data- Sparkling wine sales

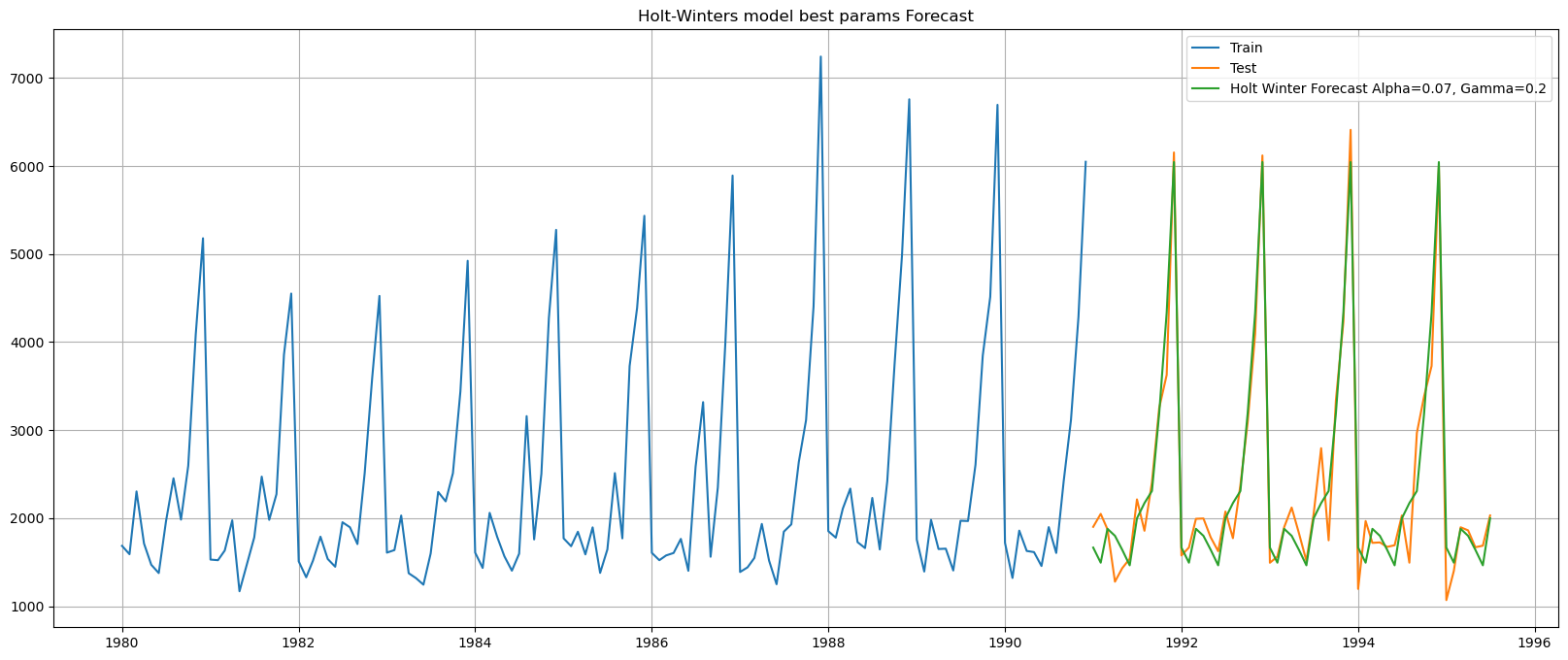


Fig.1.20. Holt Winters smoothing forecast of test data- Sparkling wine sales - optimized for lowest RMSE

**Observations:**

* + - RMSE is the lowest for alpha=0.07, and gamma=0.2, irrespective of beta value
    - We noticed that the given series has no trend, and hence the results obtained are consistent with the assumptions

1. **Check for the stationarity of the data on which the model is being built on using appropriate statistical tests and also mention the hypothesis for the statistical test. If the data is found to be non-stationary, take appropriate steps to make it stationary. Check the new data for stationarity and comment. Note: Stationarity should be checked at alpha = 0.05.**

The stationarity of the data can be ascertained by the Dickey-Fuller test. The Null and alternate hypothesis are as follows:

* **H0: The series is non-stationary**
* **Ha: The series is stationary**

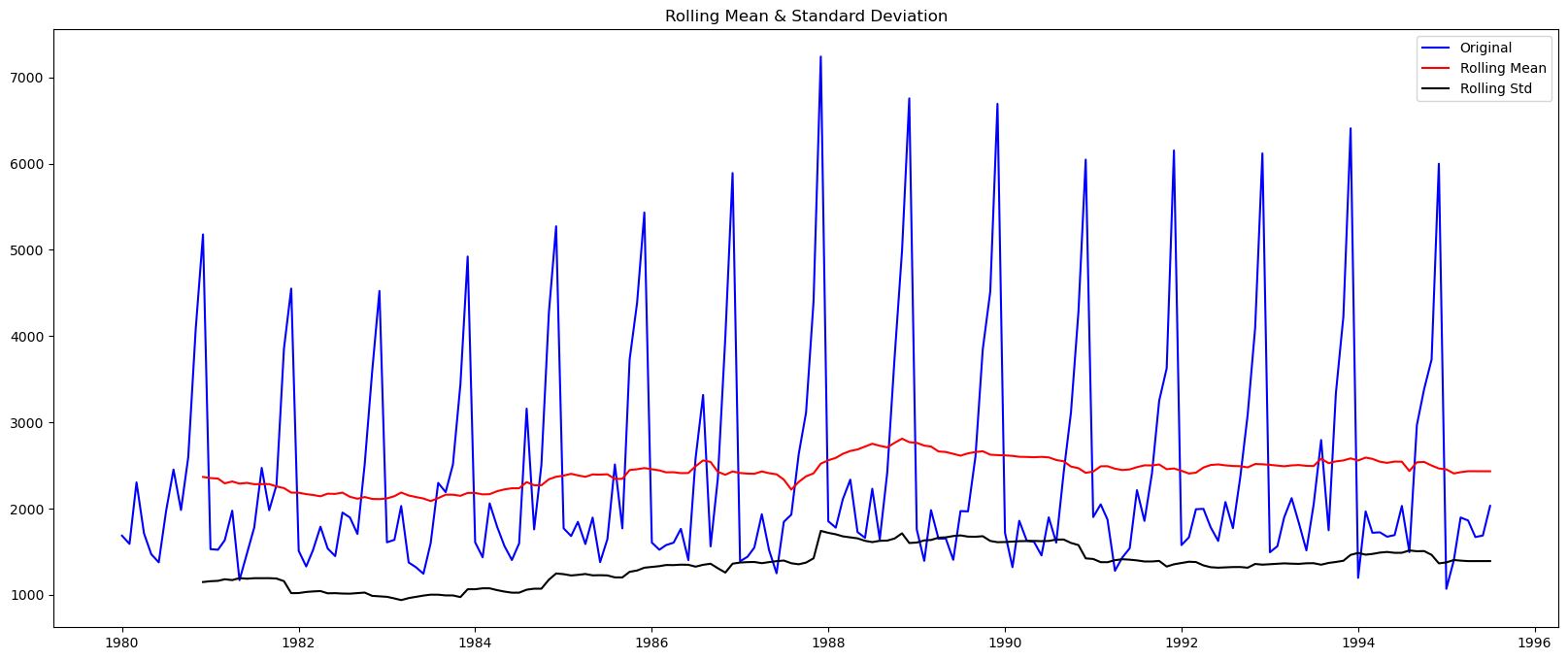
****

Fig.1.21. Sparkling dataset -Stationarity test rolling mean and Standard deviation plots

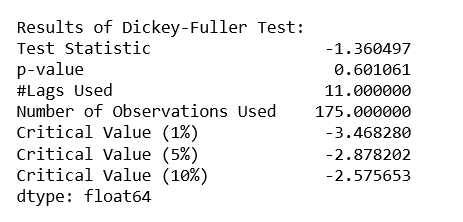


Fig.1.22. Dickey Fuller Test results- Sparkling Dataset

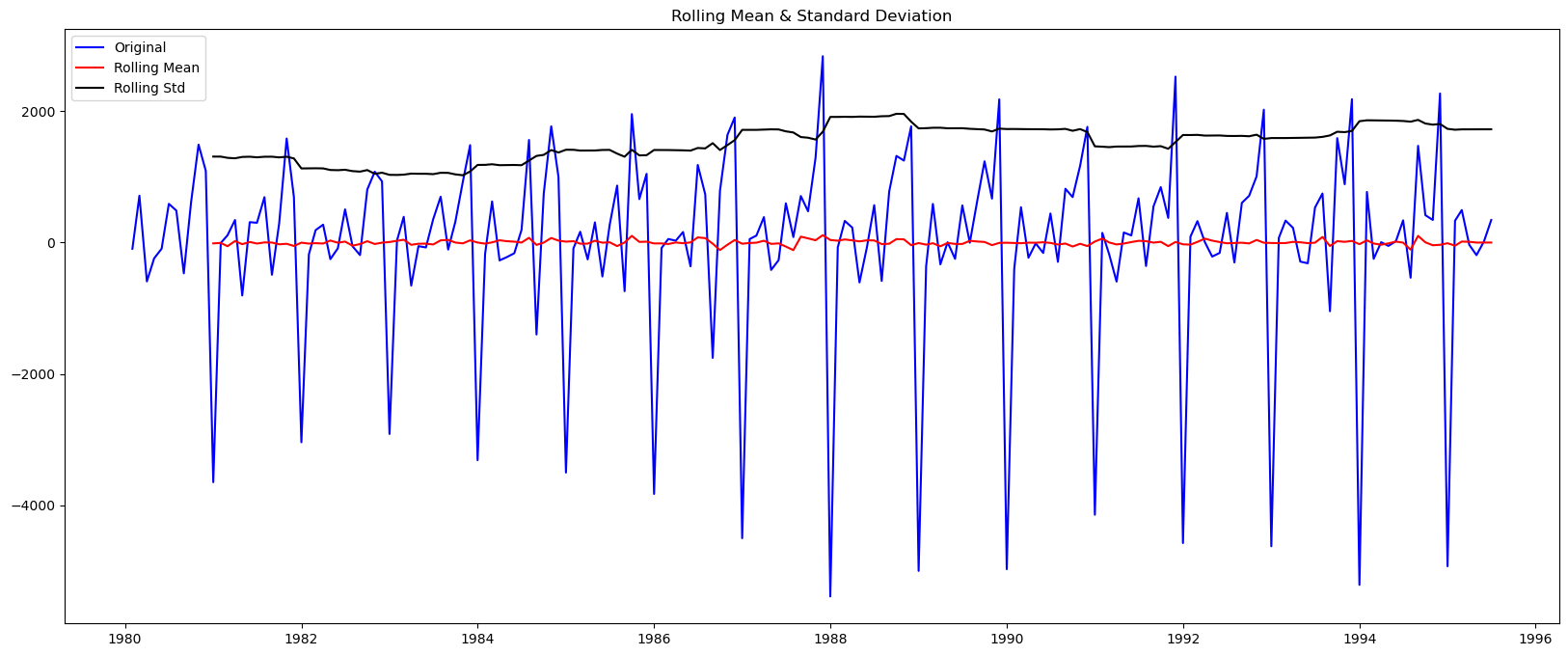
****

Fig.1.23. Differenced series -Stationarity test rolling mean and Standard deviation plots

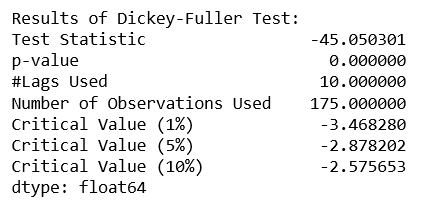


Fig.1.24. Dickey Fuller Test results- Differenced series

**Observations:**

* The given series was originally non- stationary, as evidenced by the Dickey Fuller test, with resulted in a p-value of 0.6
* After performing a first order differencing, stationarity was established. The Dickey fuller test on the differenced series resulted in a p-value of 0.0, which is less than the critical value of 0.05.

1. **Build an automated version of the ARIMA/SARIMA model in which the parameters are selected using the lowest Akaike Information Criteria (AIC) on the training data and evaluate this model on the test data using RMSE.**
   1. **ARIMA model**

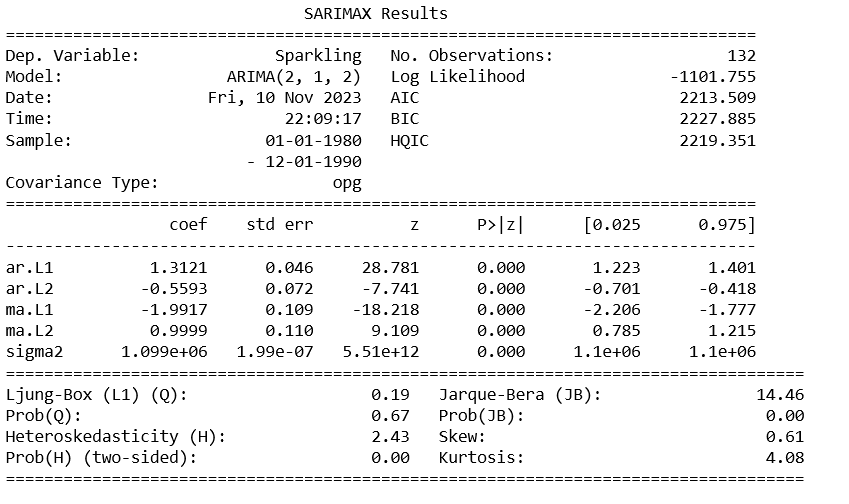


Fig.1.25. ARIMA results Summary- Sparkling Dataset

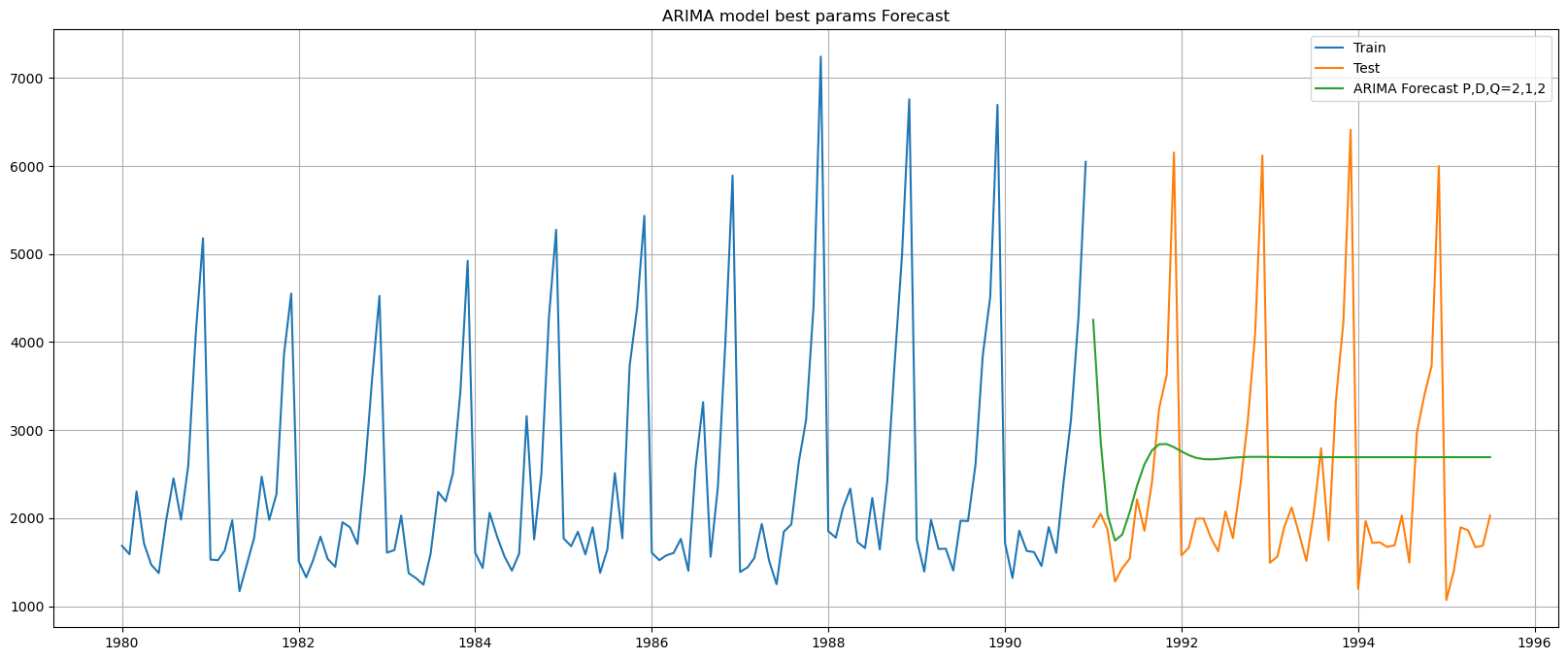


Fig.1.26. ARIMA model forecast on test data- Sparkling Dataset

**Observations:**

* + Lowest AIC obtained for (p,d,q)=(2,1,2)
  + This is consistent with the d=1 obtained during stationarity check
  1. **SARIMA model**

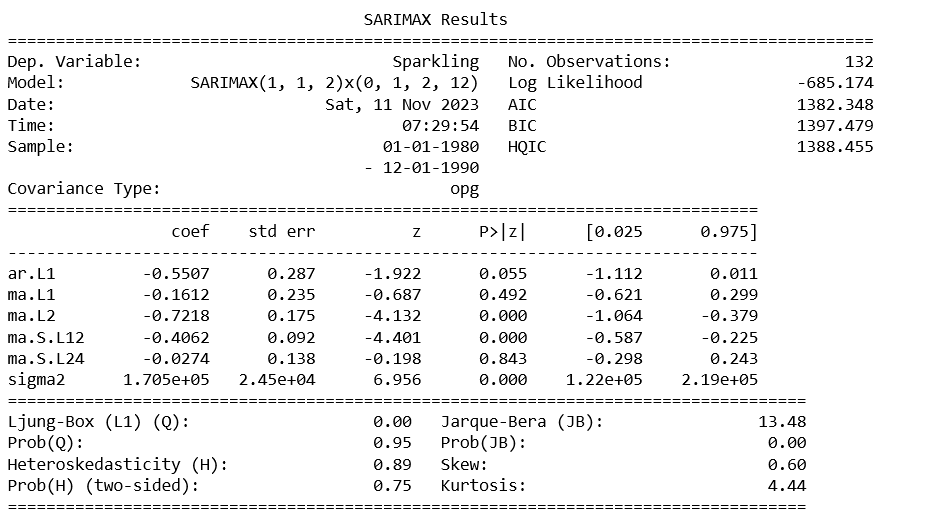


Fig.1.27. SARIMA results Summary- Sparkling Dataset

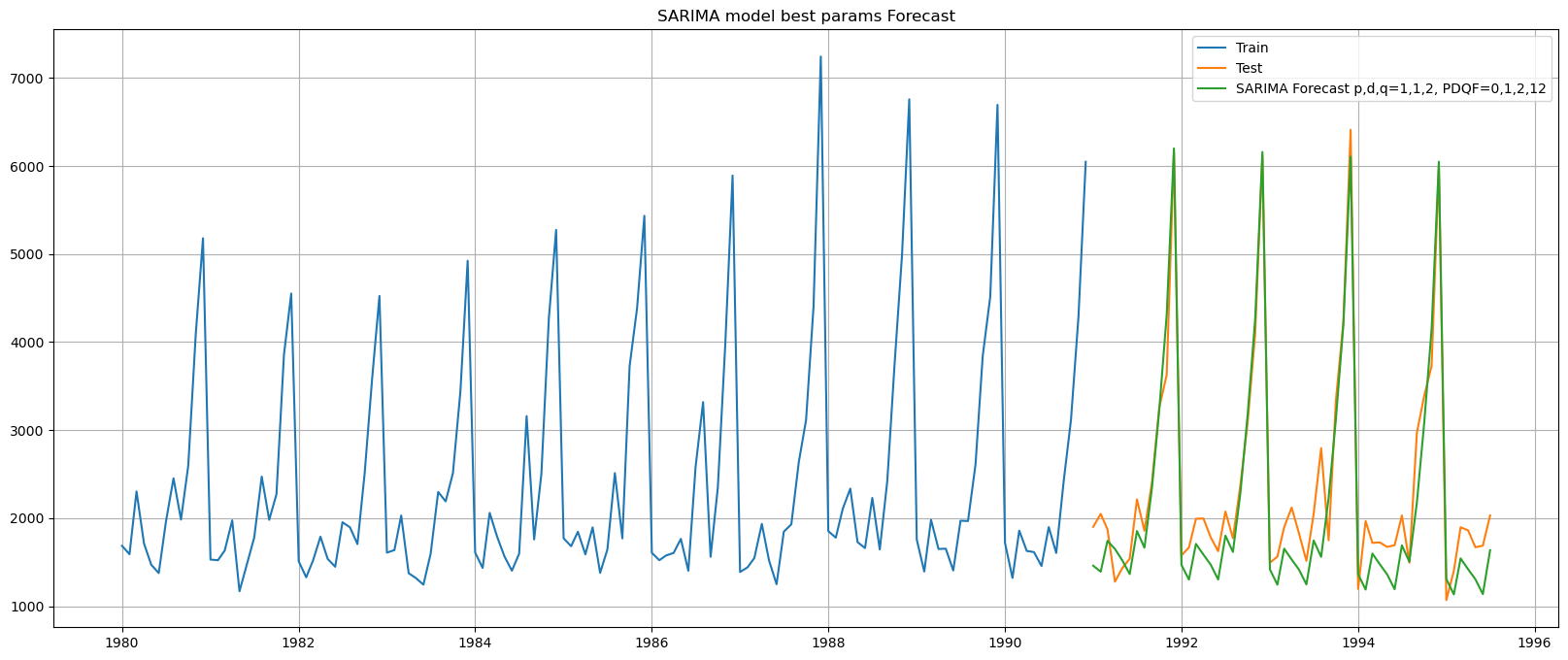
****

Fig.1.28. SARIMA model forecast on test data- Sparkling Dataset

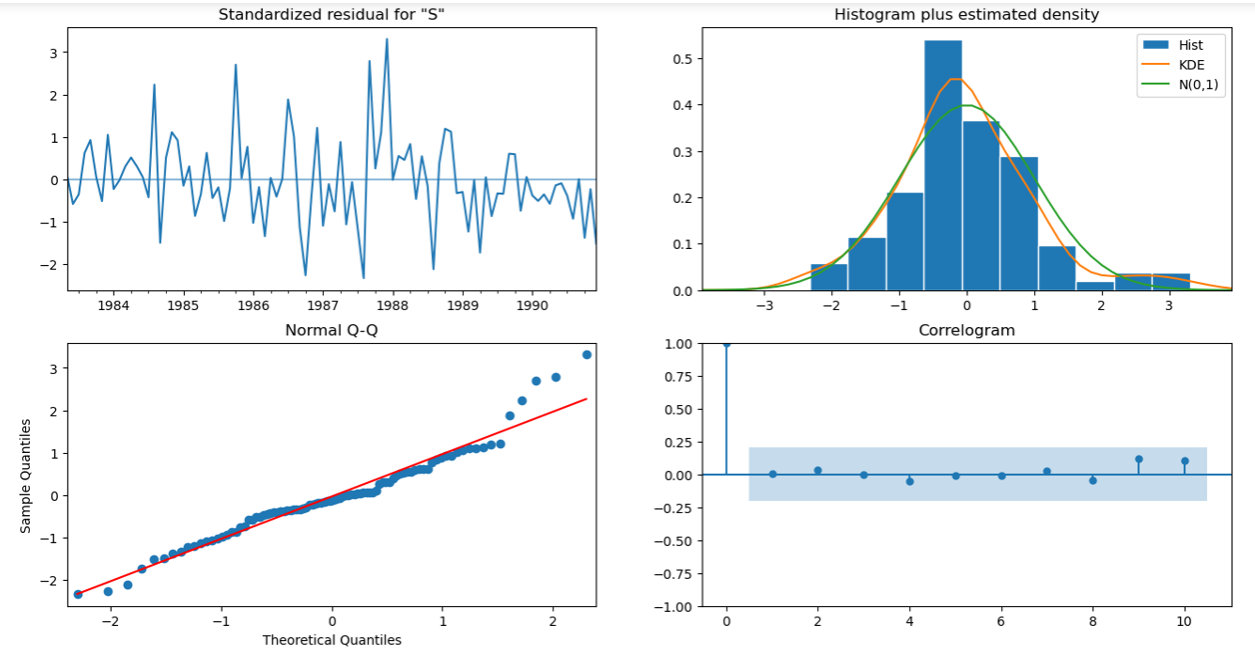


Fig.1.29. Diagnostics and Correlogram- SARIMA Model

**Observations:**

* + Lowest AIC obtained for (p,d,q)x(P,D,Q,F)=(1,1,2)x(0,1,2,12)
  + This is consistent with the d=1 obtained during stationarity check

1. **Build a table (create a data frame) with all the models built along with their corresponding parameters and the respective RMSE values on the test data.**

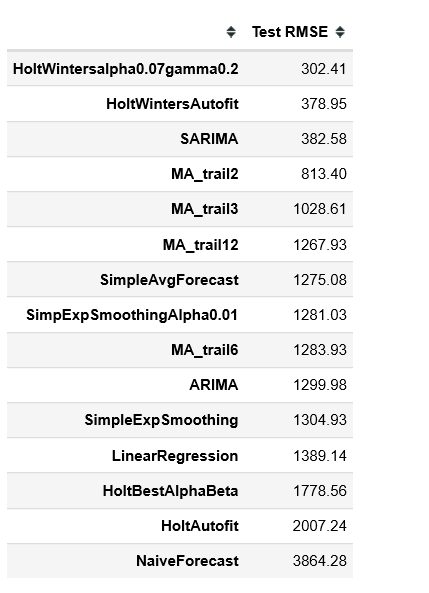


Fig.1.30. Sparkling Dataset model Results- Test RMSE

Observations:

* Lowest RMSE is obtained for Holt winters model having alpha=0.07 and gamma=0.2
* And SARIMA model for the above said parameters

1. **Based on the model-building exercise, build the most optimum model(s) on the complete data and predict 12 months into the future with appropriate confidence intervals/bands.**

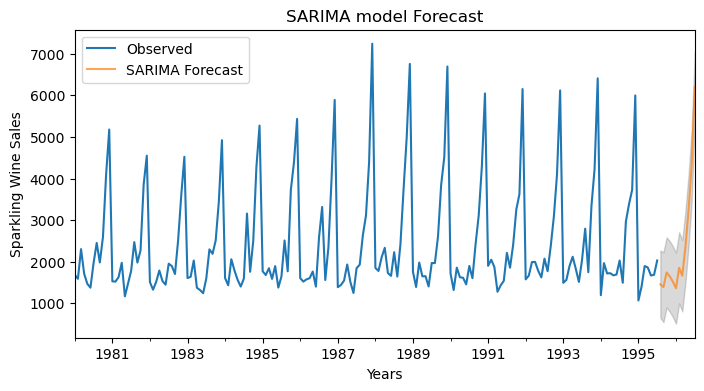


Fig.1.31. SARIMA Model forecast for next 12 months- Sparkling Dataset

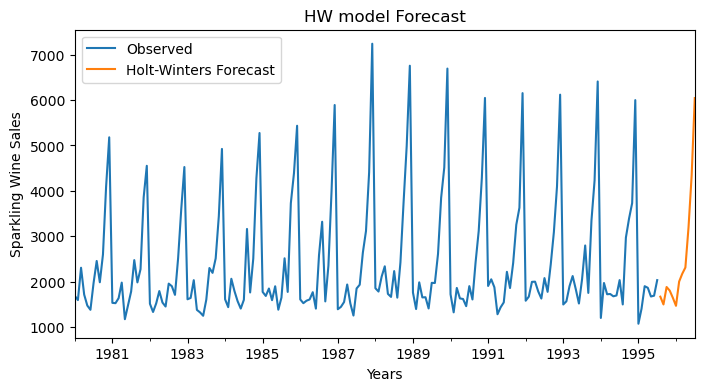


Fig.1.32. Holt Winters Model forecast for next 12 months- Sparkling Dataset

1. **Comment on the model thus built and report your findings and suggest the measures that the company should be taking for future sales.**

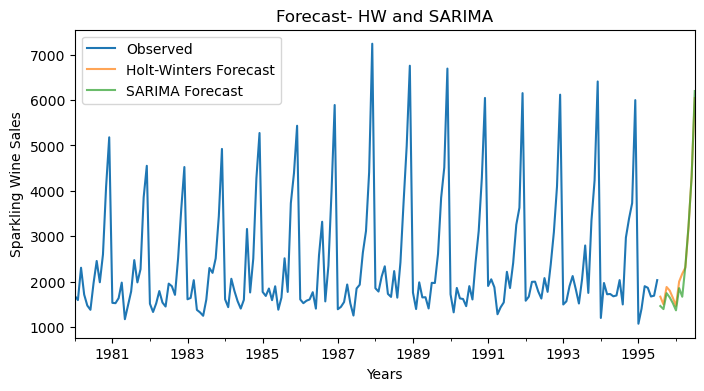


Fig.1.33. Forecast for 12 months- SARIMA and HW

**Observations:**

* The wine sales peaks during the months of November and december, probably due to the holiday season.
* The sales data does not exhibit any trend
* The forecast replicates the existing seasonality

**Insights:**

* The seasonality component of sales can be capitalized, and can try to push sales in the peak months
* The trend component needs improvement. The company can adopt different marketing strategies by customer segmentation in order to increase the overall trend